THE LANCET
Microbe

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Townsend, Hayley B Hassler, Zheng Wang, et al. The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. *Lancet Microbe* 2021; published online Oct 1. https://doi.org/10.1016/S2666-5247(21)00219-6.

The durability of immunity against reinfection by SARS-CoV-2

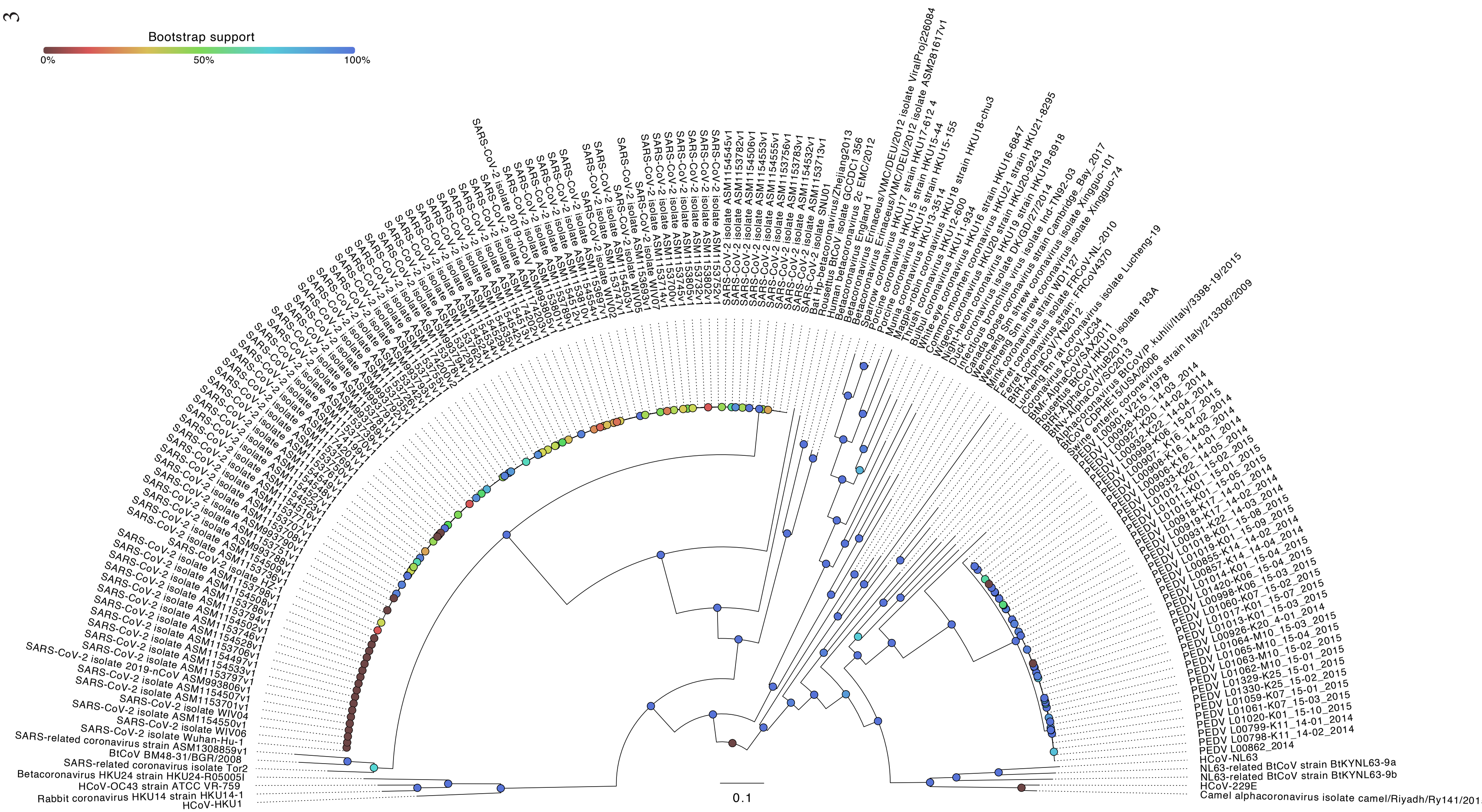
Supplementary materials

Contents

Investigators

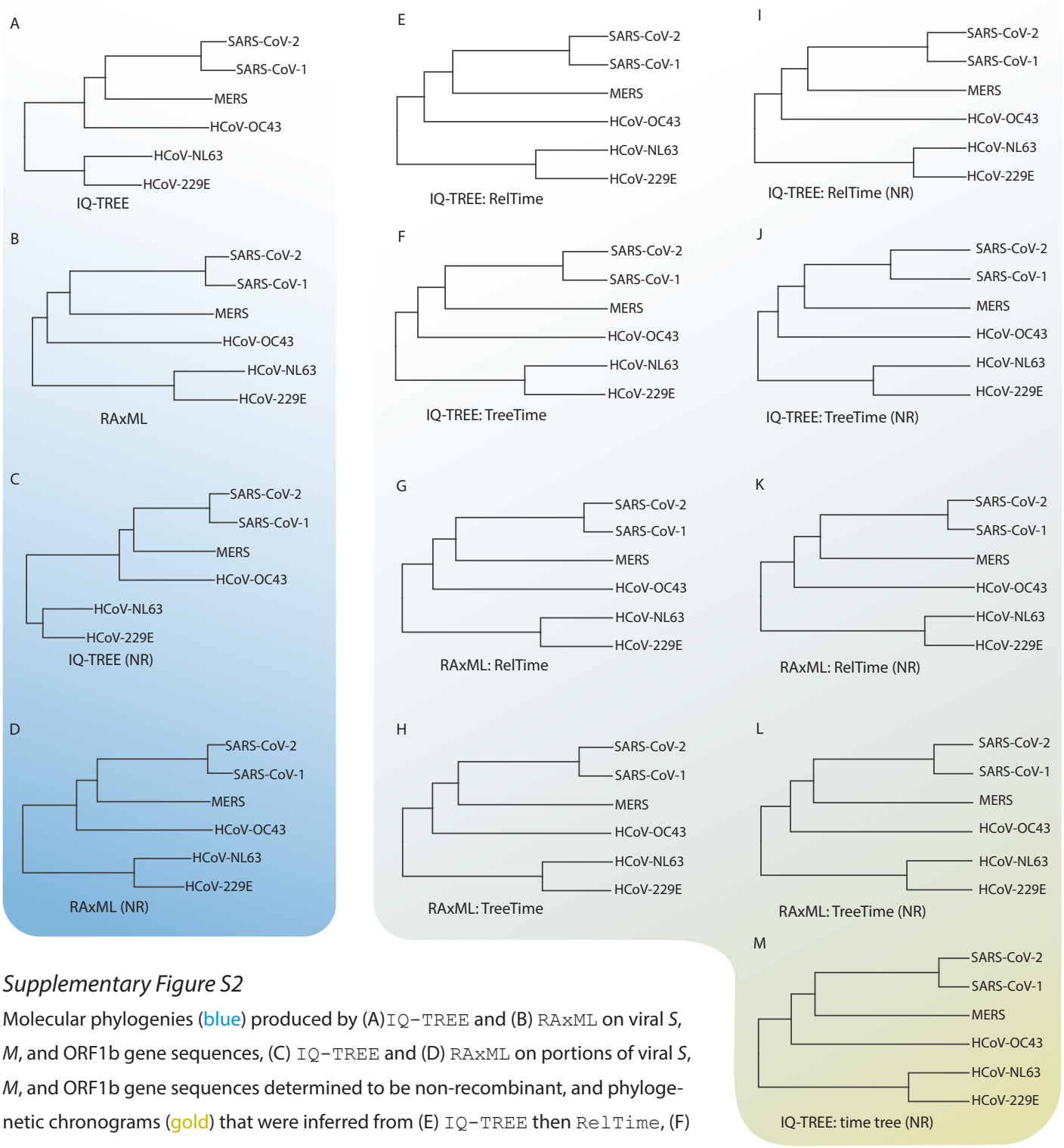
Jeffrey P. Townsend, PhD[1,2,3,4]
Hayley B. Hassler, BS[1]
Zheng Wang, PhD[1]
Sayaka Miura, PhD[5]
Jaiveer Singh[6]
Sudhir Kumar, PhD[6]
Nancy H. Ruddle, PhD[7]
Alison P. Galvani, PhD[2,7,8]
Alex Dornburg, PhD[9]

[1]Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut 06510, USA
[2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06525, USA
[3]Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06511, USA
[4]Program in Microbiology, Yale University, New Haven, Connecticut 06511, USA
[5]Institute for Genomics and Evolutionary Medicine, and Department of Biology, Temple University, Philadelphia, Pennsylvania 19122, USA
[6]Yale College, New Haven, Connecticut
[7]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut 06520
[8]Center for Infectious Disease Modeling and Analysis, Department of Epidemiology of Microbial Disease, Yale School of Public Health, New Haven, Connecticut 06525
[9]Department of Bioinformatics and Genomics, University of North Carolina, Charlotte, NC 28223



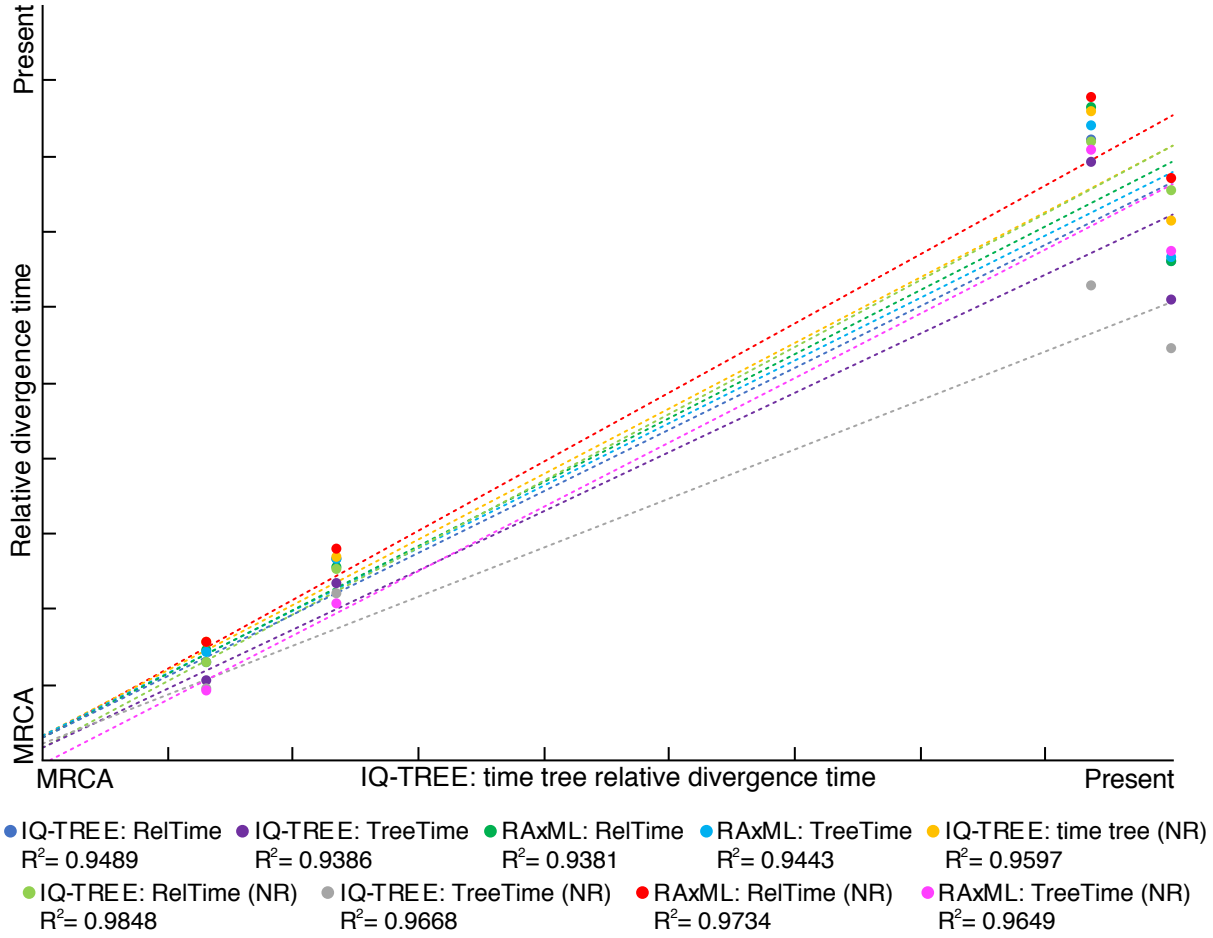
Supplementary Figure S1

Molecular phylogeny inferred using maximum likelihood[7] on the aligned sequences of *S*, *M*, and ORF1b genes for 58 *Alphacoronavirus*, 105 *Betacoronavirus*, 11 *Deltacoronavirus*, and three *Gammacoronavirus* taxa. Support for internodes was assessed by nonparametric bootstrap of sequence data (warm colors e.g. red or orange or yellow: low bootstrap support; cool colors e.g. dark blue: high bootstrap support).



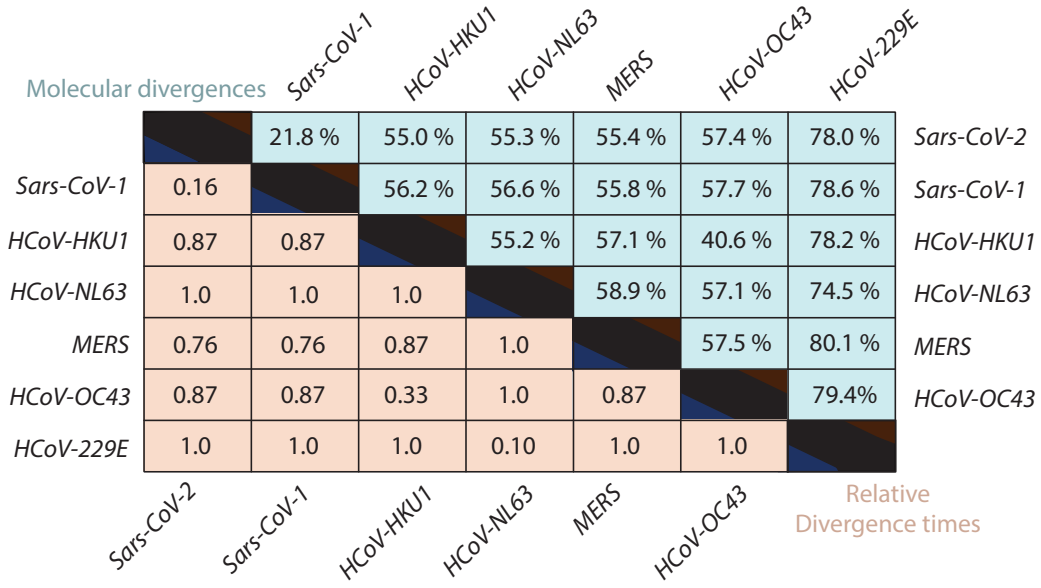
Supplementary Figure S2

Molecular phylogenies (blue) produced by (A)`IQ-TREE` and (B) `RAxML` on viral *S*, *M*, and ORF1b gene sequences, (C) `IQ-TREE` and (D) `RAxML` on portions of viral *S*, *M*, and ORF1b gene sequences determined to be non-recombinant, and phylogenetic chronograms (gold) that were inferred from (E) `IQ-TREE` then `RelTime`, (F) `IQ-TREE` then TreeTime, (G) `RAxML` then `RelTime`, (H) `RAxML` then `TreeTime`, (I) `IQ-TREE` then `RelTime` on non-recombinant sequences, (J) `IQ-TREE` then `TreeTime` on non-recombinant sequences, (K) `RAxML` then `RelTime` on non-recombinant sequence (L) `RAxML` then `TreeTime` on non-recombinant sequences, and (M) `IQ-TREE` on non-recombinant sequences, conditioned on the evolutionary divergences of human-infecting coronaviruses. Bootstrap support was 100% for all nodes of these phylogenies. NR: Based on viral gene sequences that were determined to be non-recombinant.



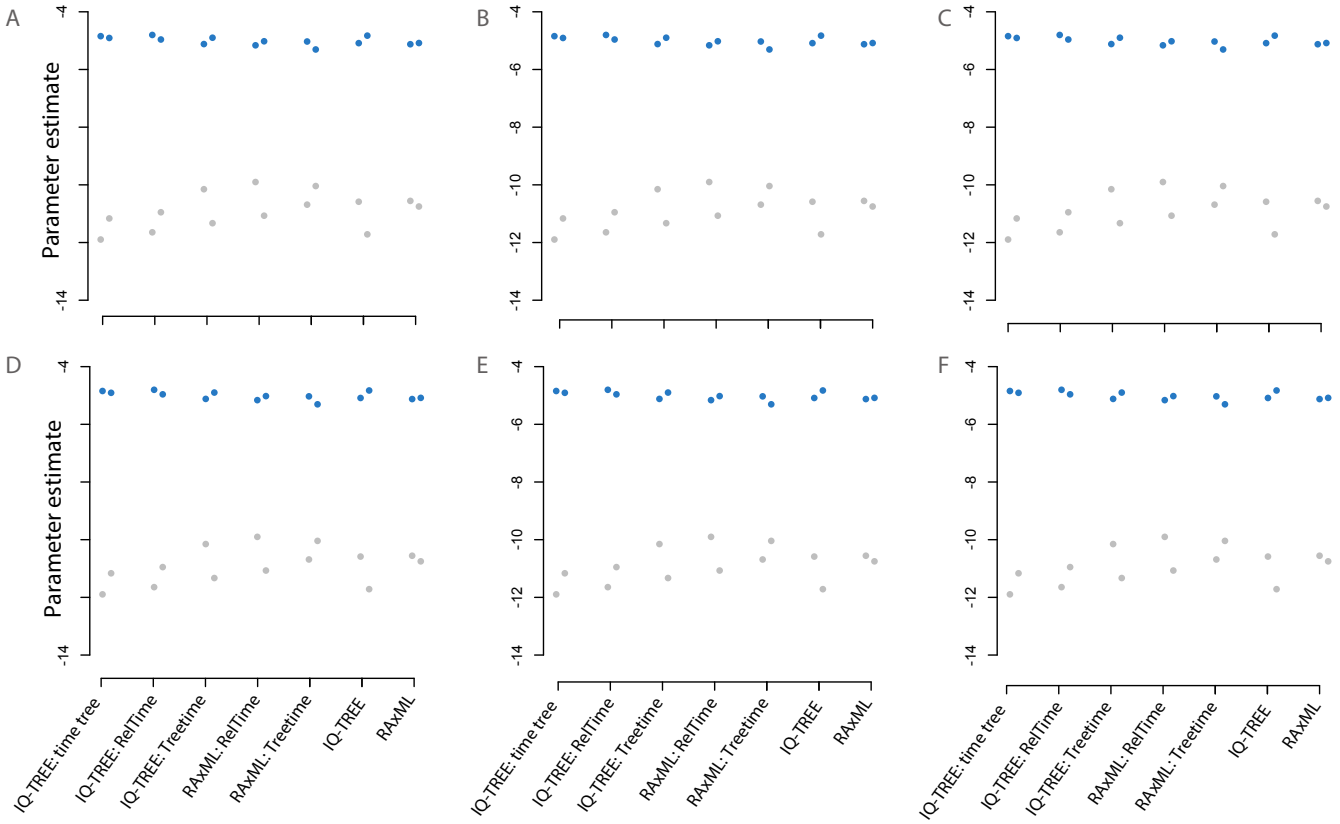
Supplementary Figure S3

Scatter plot of the node ages estimated by phylogenetic chronograms derived from IQ-TREE and RAxML molecular phylogenies compared to the IQ-TREE estimated divergence times. NR: Based on viral gene sequences that were determined to be non-recombinant.



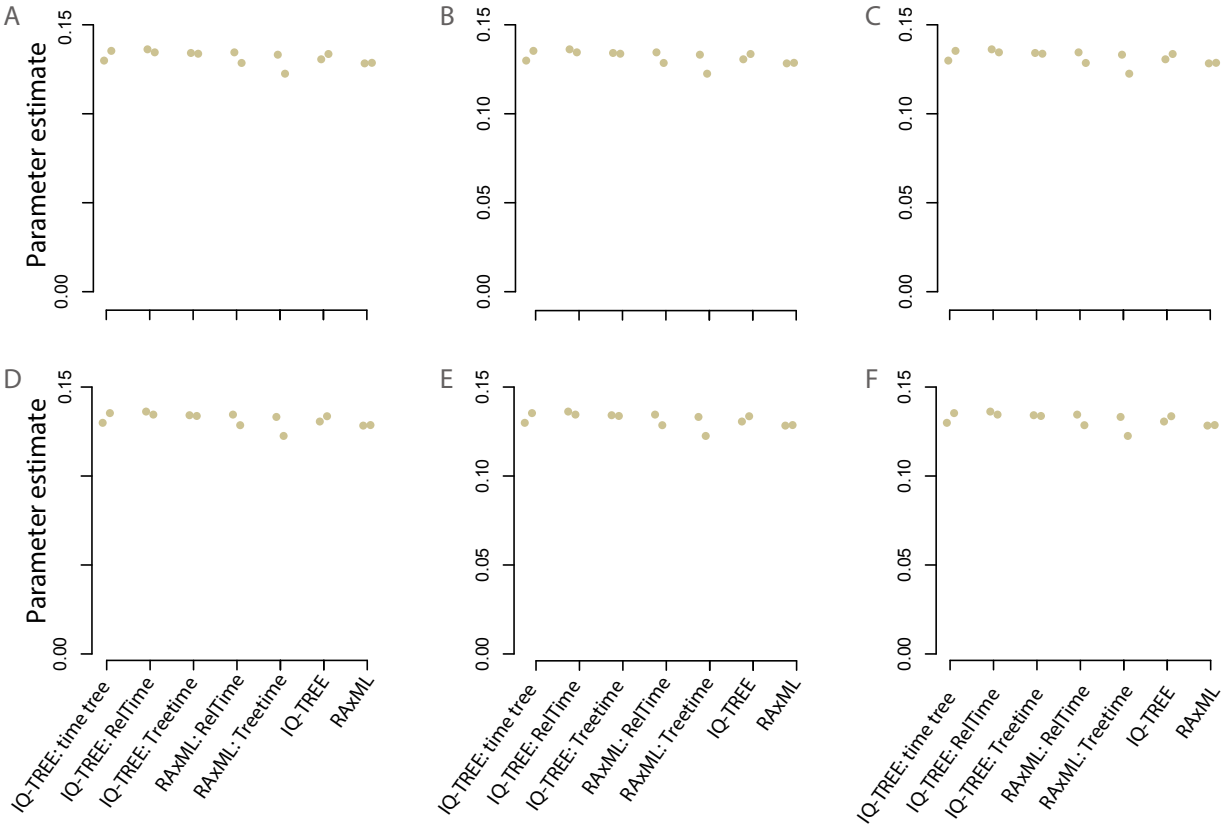
Supplementary Figure S4

Viral percent sequence divergences (blue; based on the alignment of the *S*, *M*, and ORF1b genes), and estimated pairwise divergence times (beige; from our `IQ-TREE` phylogenetic chronogram analysis, with time calibrated relative to the MRCA at 1.0).



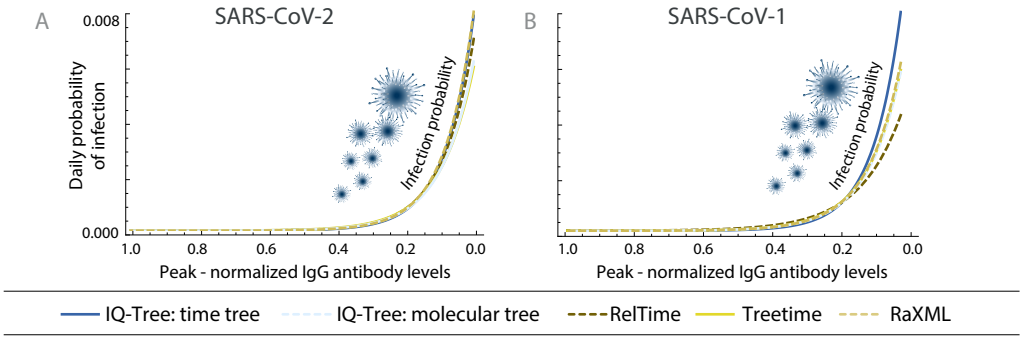
Supplementary Figure S5

Evaluation of the sensitivity of the `Rphylopars`estimate based on A) anti-N IgG; B) anti-S IgG; C) anti-whole-virus IgG; D) an alternate data source for SARS-CoV-1 anti-N IgG; E) an alternate data source for MERS-CoV anti-S IgG; and F) an alternate data source for anti-whole virus IgG to the slope (*b*, blue) and intercept (*a*, gray) parameters for the logistic regression function, as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference.



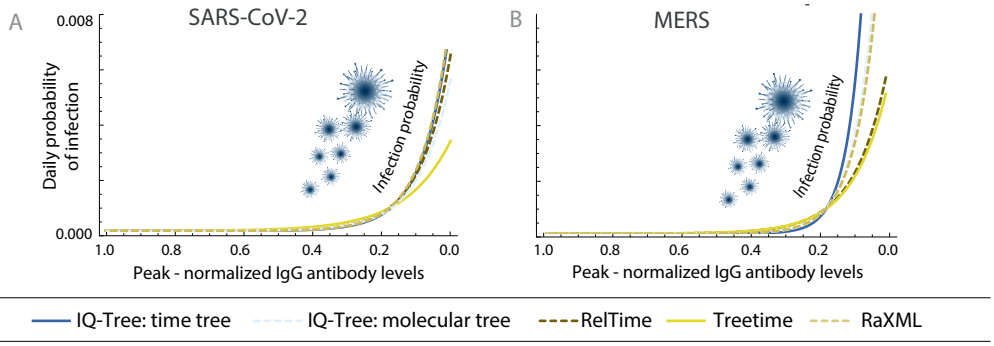
Supplementary Figure S6

Evaluation of the sensitivity of the `Rphylopars`estimate based on A) anti-N IgG; B) anti-S IgG; C) anti-whole-virus IgG; D) an alternate data source for SARS-CoV-1 anti-N IgG; E) an alternate data source for MERS-CoV anti-S IgG; and F) an alternate data source for anti-whole virus IgG to the baseline parameter for the logistic regression function, as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference.



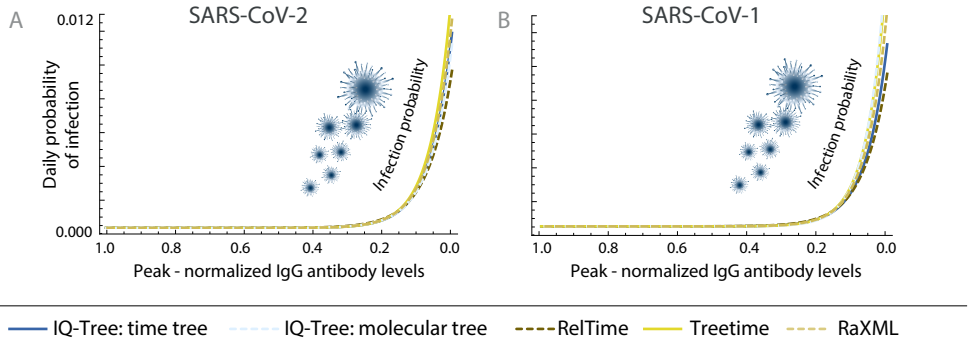
Supplementary Figure S7

Evaluation of the sensitivity of the `Rphylopars`-based estimation of the relationship between the daily infection probability and the peak normalized anti-N protein Immunoglobulin G antibody levels for A) SARS-CoV-2 & B) SARS CoV-1 as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference. Anti-N IgG data was not available for MERS-CoV.



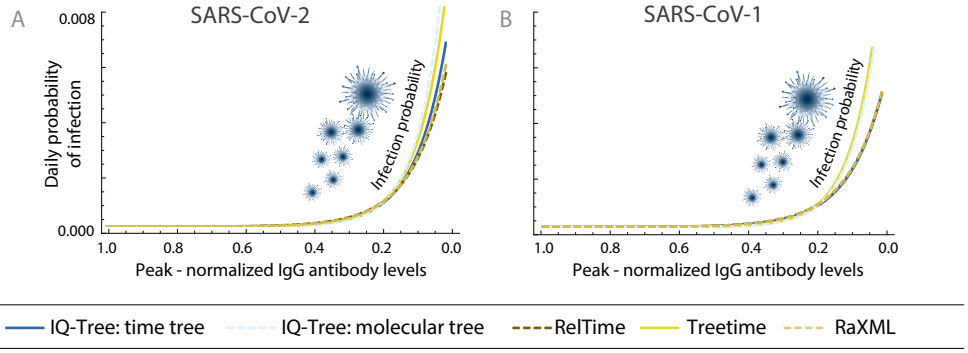
Supplementary Figure S8

Evaluation of the sensitivity of the `Rphylopars`-based estimation of the relationship between the daily infection probability and the peak normalized anti-S-protein Immunoglobulin G antibody levels for A) SARS-CoV-2 & B) MERS-CoV as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference. Anti-S IgG data was not available for SARS-CoV-1.



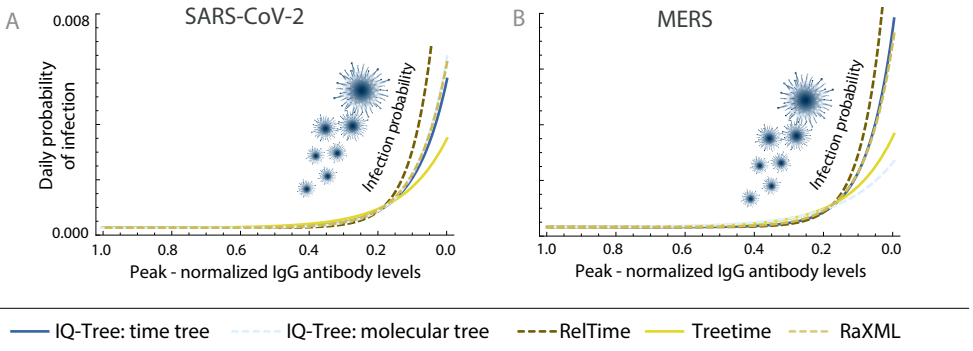
Supplementary Figure S9

Evaluation of the sensitivity of the `Rphylopars`-based estimation of the relationship between the daily infection probability and the peak normalized anti-whole virus Immunoglobulin G antibody levels for A) SARS-CoV-2 & B) SARS CoV-1 as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference. Anti whole-virus IgG data was not available for MERS-CoV.



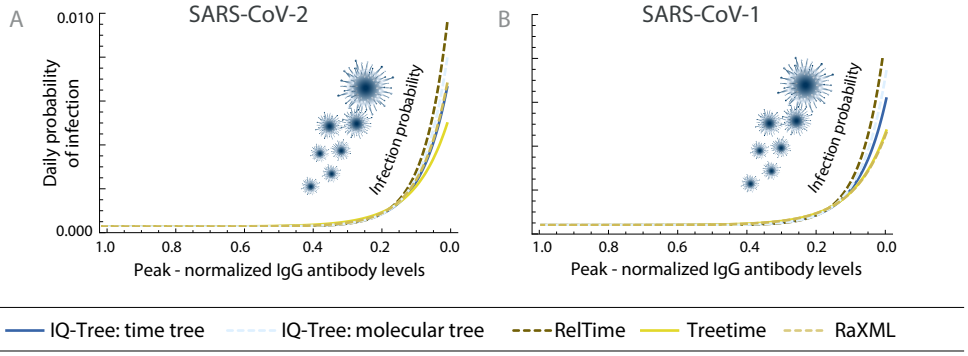
Supplementary Figure S10

Evaluation of the sensitivity of the `Rphylopars`-based estimation of the relationship between the daily infection probability and the peak normalized anti-N protein Immunoglobulin G antibody levels using an alternate data source for SARS-CoV-1 as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference. Anti-N IgG data was not available for MERS-CoV.



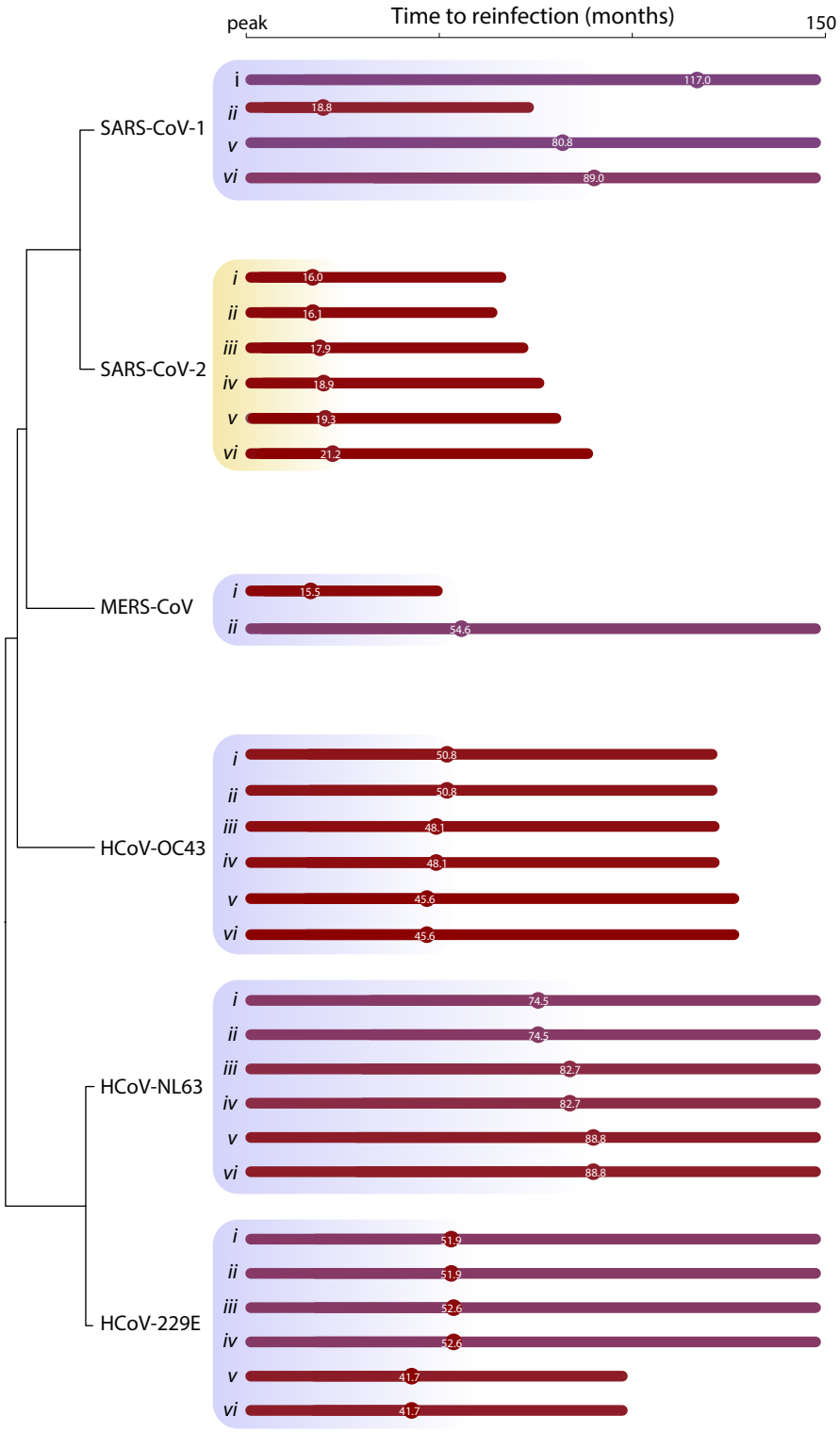
Supplementary Figure S11

Evaluation of the sensitivity of the `Rphylopars`-based estimation of the relationship between the daily infection probability and the peak normalized anti-S-protein Immunoglobulin G antibody levels A) SARS-CoV-2 & B) MERS-CoV using an alternate data source for MERS-CoV as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference. Anti-S IgG data was not available for SARS-CoV-1.



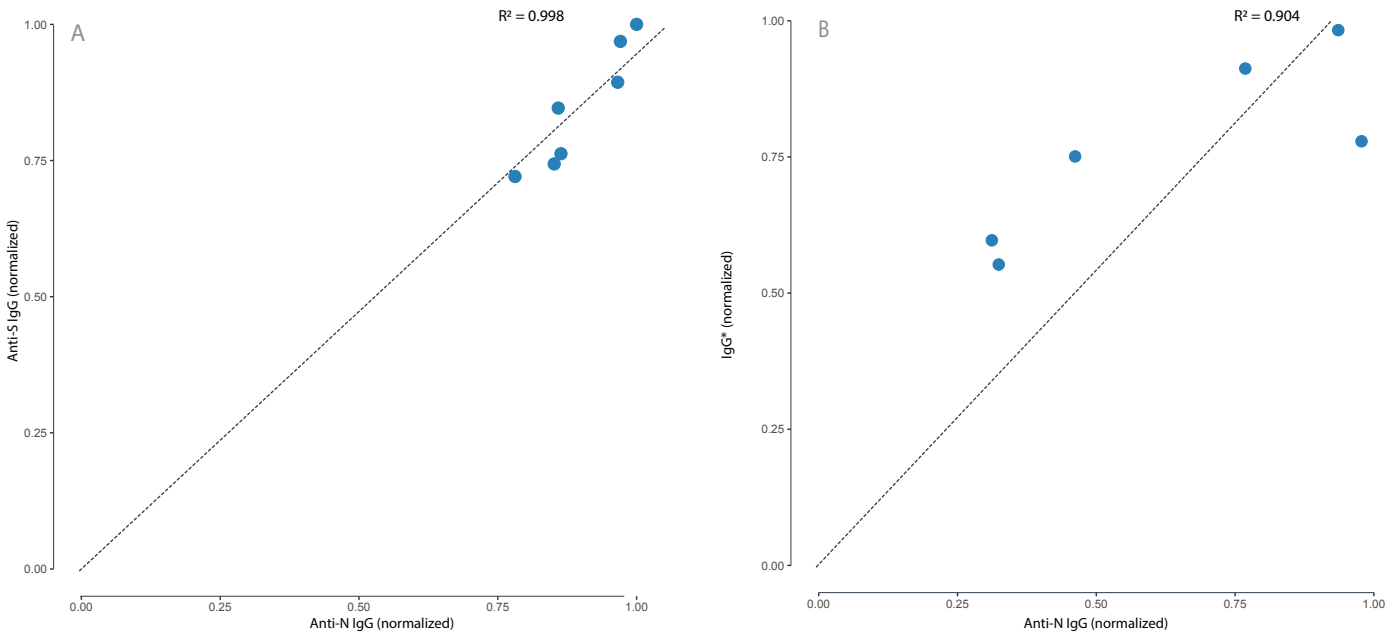
Supplementary Figure S12

Evaluation of the sensitivity of the `Rphylopars`-based estimation of the relationship between the daily infection probability and the peak normalized anti-whole-virus Immunoglobulin G antibody levels from an alternate data source for SARS-CoV-1 for A) SARS-CoV-2 & B) SARS CoV-1 as they depend on molecular divergence phylogenies or phylogenetic chronograms generated via several common approaches to phylogenetic inference. Anti while-virus IgG data was not available for MERS-CoV.



Supplementary Figure S13

Evaluation of the sensitivity of the times to reinfection for each coronavirus resulting from analyses of datasets i–vi (detailed in the Fig. 1 legend), in days from peak antibody level at 3 months.



Supplementary Figure S14

Results of a linear regression demonstrating the relationship between (A) normalized anti-N IgG and normalized anti-S IgG levels ($P = 6.63 \times 10^{-9}$) and (B) normalized anti-N and normalized anti-virus lysate IgG levels ($P = 0.000623$) at the same time points. Dotted line indicates the expectation of linear model fit through the origin.